The Dissertation Committee for Pulak Goswami
certifies that this is the approved version of the following dissertation:

Recovering the Payoff Structure of a Utility Maximizing Agent

Committee:

Gordan Žitković, Supervisor

Kasper Larsen

Nataša Pavlović

Mihai Sirbu

Recovering the Payoff Structure of a Utility Maximizing Agent

by

Pulak Goswami, B.S. In Math.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2016

Dedicated to my parents, Anil and Rekha Goswami.

Acknowledgments

I wish to acknowledge my Adviser, Gordan Žitković, as well as my colleagues without whom this work would not have been possible.

Recovering the Payoff Structure of a Utility Maximizing Agent

Publication No. ____________________

Pulak Goswami, Ph.D.
The University of Texas at Austin, 2016

Supervisor: Gordan Žitković

Any agent with access to information that is not available to the market at large is considered an 'insider'. It is possible to interpret the effect of this private information as change in the insider's probability measure. In the case of exponential utility, logarithm of the Radon-Nikodym derivative for the change in measure will appear as a random endowment in the objective the insider would maximize with respect to the original measure. The goal of this paper is to find conditions under which it is possible to recover the structure of this random endowment given only a single trajectory of his/her wealth. To do this, it is assumed that the random endowment is a function of the terminal value of the state variable and that the market is complete.

Table of Contents

Chapter 1

Introduction

An 'insider' is any economic agent who has privileged information about the underlying asset in the market. That is to say, he has knowledge about the underlying that is not available to the market at large. This extra knowledge should, in any market model, provide additional economic value to the agent. How may he/she optimize this value? This question has been approached in two basic frameworks: the auction or equilibrium framework and filtration-enlargement framework.

The auction framework was first studied by Kyle [9] and later generalized by Back [2]. Kyle [9] considered a market consisting of three types of agents: the insider, the market makers and the noisy traders. There is one underlying asset $v \sim \mathcal{N}(\alpha, \phi)$ and the insider knows its terminal value: $\tilde{v}$. Trading occurs only once and at this time the three types of agents engage in trading the asset. This is done by placing (possibly negative) orders for the asset and it happens in two steps. First, the insider and noisy traders, having no knowledge of each other, place their orders x and u respectively. It is assumed that the noisy traders place orders distributed as $\mathcal{N}(0, \sigma)$ regardless of the current price of the asset. Second, the market makers determine the price p at which they will trade the quantity to clear the market. The market makers are assumed to be 'rational traders' in that they set the price such that:

$$P(x+u) = \mathbb{E}\left[\tilde{v} | x+u\right] \tag{1.1}$$

In setting his/her order x, the insider knows the market makers will obey the rule (1.1) above. Therefore, he/she is aware of the effect his/her trade will have on the price. Knowing this, the insider places an order x that maximizes his profit $\tilde{\pi} = (\tilde{v} - p)x$. Therefore, the trading price is an equilibrium between the insider's order x and the price p that maximizes $\tilde{\pi}$. In other words, it is any pair (x, p) that satisfies

$$\mathbb{E}\left[\tilde{\pi}(x,p) | \tilde{v} = v\right] \geq \mathbb{E}\left[\tilde{\pi}(x',p) | \tilde{v} = v\right] \tag{1.2}$$

for any alternate x' and any true value v as well as the market clearing rule (1.1).

In this setup Kyle shows that the equilibrium is linear:

$$X(\tilde{v}) = \beta(\tilde{v} - p_0) \quad \text{and} \quad P(x+u) = p_0 + \lambda(x+u)$$

where $\beta = \sigma/\phi$ and $\lambda = 2\phi/\sigma$ are constants that depend on the known variance of the orders placed by noisy trades, the order u, and the value of the underlying $\tilde{v}$. This result is extended to a multi-auction model where the same auction takes place multiple times at regular intervals. Taking the time between these auctions to be arbitrarily small, the model is generalized to continuous time. In this limit the equilibrium is described as:

$$dX_t = \frac{\tilde{v} - P_t}{\lambda(1-t)} dt \quad \text{and} \quad P_t = \alpha + \lambda u_t \tag{1.3}$$

Where α, λ are continuous time analogs of the one step model parameters, and u_t is a Brownian Motion that represents the orders placed by noisy traders. Back extends this model to allow for a general distribution of the asset value [9].

The key feature of these auction type models is that they describe the price of the underlying as a function of the insider's orders x. Despite having knowledge of the true value of the underlying, the insider is not able to make infinite profit. This is because the magnitude of his/her order has an inverse effect on the trading price P_t. Note also that it is crucial that the market makers observes $x_t + u_t$ together and not the insider's orders x_t directly. Knowing x_t exactly, the market makers would be able to recover $\tilde{v}$ exactly by numerically inverting the integral (1.3) described above.

Another approach [10] seeks to model the insider within the Merton framework as an agent with an enlarged filtration. In contrast to the auction model, here the price of the underlying asset is exogenous. That is, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathcal{F} = \{\mathcal{F}_t\}_{0 \leq t \leq 1}$ generated by the Brownian Motion satisfying the usual conditions of completeness and right-continuity, the prices are described by the following stochastic differential equations:

$$dP_t^0 = P_t^0 r_t dt \tag{1.4}$$

$$dP_t^i = P_t^i \left[b_t^i dt + \sum_{j=1}^{d} \sigma_t^{ij} dW_t^j \right] \tag{1.5}$$

The wealth of an agent is given by $X_t^\pi = \int_0^t \pi_s \cdot dP_s$, where x denotes the initial wealth. The insider, like any rational agent, solves a utility maximization problem:

$$V(x) = \sup_{\pi \in \mathcal{A}} \mathbb{E}\left[U(X_1^\pi)\right] \tag{1.6}$$

for some set $\mathcal{A}$ of admission portfolio processes. This problem was first studied by Pikovsky and Karatasas [10] in the complete market setting and then extended to the incomplete setting by Amendinger [1]. Both of these papers take $U(x) = \log(x)$.

Similar to Kyle's model, the insider here has some knowledge about the terminal price of the underlying from the beginning. In this context, this would mean that the insider has knowledge of an $\mathcal{F}_1$ measurable random variable L at time $t = 0$. He/she, then, has the ability to choose from portfolios adapted to the enlarged filtration $\mathcal{G}_t$ given by:

$$\mathcal{G}_t = \mathcal{F}_t \vee \sigma(L)$$

To understand the effect of this additional information both [10] and [1] use (initial) filtration-enlargement theory, most of which was brought forth by Jacod [5]. Jacod works under the following assumption on the random variable L:

Assumption 1. *There exists a σ-finite measure η on $(\Omega, \mathcal{F})$, such that for any $t \in [0,1)$ the regular conditional probability $\mathbb{P}\left[L \in dx | \mathcal{F}_t\right]$ is absolutely continuous with respect to η for $\mathbb{P}$-a.e. $\omega \in \Omega$.*

This assumption allows for natural choices of L such as $L = W_1$ or $L = W_1 + \epsilon$, with ϵ representing independent noise, this yields the following remarkable result:

Theorem 1 (Jacod [5]). *There exists a $\mathcal{G}_t$-measurable function $(\omega, l, t)^i \to (k_t^l(\omega))^i$, for each $i = 1, ..., d$, such that:*

- $\int_0^t |(k_s^L)^i| dt < \infty \quad \mathbb{P}$*-a.s. for all* $t \in [0, 1)$.
- *Each W_t^i is a $\mathcal{G}_t$ semimartingale with the following canonical decomposition:*

$$\tilde{W}_t^i = W_t^i - \int_0^t (k_s^L)^i ds, \quad t \in [0, 1) \tag{1.7}$$

Here, $k_t^l(\omega)$ being $\mathcal{G}_t$ measurable means that for any fixed l, the function k_t^l is measurable with respect to $\mathcal{G}_t$. A notable example that resembles the privileged information in Kyle [9] is $L = W_1$ which yields

$$\hat{W}_t = W_t - \int_0^t \frac{W_1 - W_s}{1 - s} ds \tag{1.8}$$

It is interesting to observe that $\hat{W}$ becomes a Brownian bridge.

The decomposition (1.7) allows us to rewrite the dynamics of the price P_t on the space $(\Omega, \mathcal{G}, \mathbb{P})$ equipped with the enlarged filtration $\mathcal{G}_t$. The risk free price remains the same, and the risky assets become:

$$dP_t^i = P_t^i \left[\left(b_t^i + (k_t^L)^i \right) dt + \sum_{j=1}^d \sigma_t^{ij} d\tilde{W}_t^j \right] \tag{1.9}$$

From the insider's point of view, first, the random variable L is realized then the price evolves as in (1.9) described above. At this point, the insider's optimization problem (1.6) is placed well within established portfolio optimization theory. Under an appropriate choice of $\mathcal{A}$, Amendinger et al. [1] give us the following explicit result comparing the insider, who has additional knowledge of L, and the 'uninformed' economic agent who does not:

Theorem 2. *For a fixed time $t \in [0,1]$, the optimal strategy of an uninformed economic agent is given by $\pi_t^{un} = \alpha_t$ and the corresponding maximal utility is given by:*

$$\mathbb{E}\left[V_t(x,\pi^{un})\right] = \log(x) + \frac{1}{2}\mathbb{E}\left[\int_0^t ||\alpha_s||^2 ds\right] \tag{1.10}$$

The optimal strategy of the insider is given by $\pi_s^{ins} = \alpha_s + k_s^L$ and the corresponding logarithmic utility is given by:

$$\mathbb{E}\left[V_t(x,\pi^{ins})\right] = \log(x) + \frac{1}{2}\mathbb{E}\left[\int_0^t ||\alpha_s||^2 ds\right] + \frac{1}{2}\mathbb{E}\left[\int_0^t ||k_s^L||^2 ds\right] \tag{1.11}$$

Amendinger et al. also show that the difference in utility between the insider and the uninformed agent can be rewritten in terms of the relative entropy $H(\mathbb{P}|\tilde{\mathbb{P}}_t) = \mathbb{E}\left[\log \frac{d\mathbb{P}}{d\tilde{\mathbb{P}}_t}\right]$, where $\tilde{\mathbb{P}}$ is the measure under which $\tilde{W}$ is a local martingale.

Unlike in the auction type models, here the price P is not effected by the insider's portfolio choice π. Therefore, the insider can be interpreted as a 'small' agent, having no ability to manipulate the price. This assumption is reasonable only if the insider's utility in (1.6) is finite. Indeed, a very large (or very negative) demand for the underlying by the insider ought to effect its price in an open market. Pikovsky and Kartsas show that even a small amount of information about the future, if not distorted by noise, gives the insider infinite wealth. Amendinger et al. show, more precisely, that the insider's gain is finite only if $H(\mathbb{P}|\tilde{\mathbb{P}}_1)$ is finite.

1.1 Recovering Information

The models above describe the optimal behavior of the insider in great detail. Both types of models relate the private information of an insider to his/her trading strategy explicitly. With these types of explicit results, it is natural to ask if it is possible to recover the insider's private information by simply observing his trading strategies. That is to say, is it possible for the 'market makers,' or 'uninformed traders,' to decipher the insider's private information if they are able to observe his/her trading strategy.

In the auction models described by Kyle and Back, it is essential that the market makers do not observe the insider's orders directly. As mentioned above, observing X_t in (1.3) for any period of time (and knowing it's structure) would allow one to infer the quantity $\tilde{v}$. Within the model however, the trading strategy X_t is obscured by the noisy traders and the market makers observe $X_t + u_t$. Nevertheless, suppose the market makers know the following:

- That the insider has knowledge of the terminal value $\tilde{v}$.
- The combined orders $X_t + u_t$, where u_t is an independent Brownian

Motion.

- The price P_t.

By knowing the pricing rule that market makers themselves place, as well as that the insider would seek to optimize his wealth, the market makers would be able to infer that the insider would place orders satisfying (1.3). In this way, the the problem of recovering $\tilde{v}$ becomes a simple filtering problem. Given the signal:

$$X_t = \int_0^t \frac{\tilde{v} - P_t}{\lambda(1-t)} dt + u_t$$

What is the best estimate for $\tilde{v}$? Such a problem fits well within the realm of filtering theory. The effect of adding such a filtering problem to the Back's model is an interesting problem, but not discussed here.

Though such a problem is tractable, it is unreasonable to presume that the nature of the privileged information, i.e. that the insider knows the terminal value $\tilde{v}$, would be available publicly. However, it is often the case that certain agents in the market have *better* knowledge of the market. The filtration-enlargement framework allows for this interpretation. The 'insider,' by knowing the value of a random variable correlated with the price, simply has a better understanding of the factors influencing the price of the underlying asset. This 'better understanding' is made explicit in the extra drift term k_t^L in (1.7).

In the filtration enlargement framework, it would be natural to take the following as input:

- The knowledge that the insider is maximizing an objective of the type (1.6).
- The insider's portfolio π_t^{ins}.
- The stock price P_t or, equivalently, the state variable W_t.

Again, the 'uninformed agent' would be able to compute his own optimal strategy $\pi_t^{\text{un}} = \alpha_t$ and be able to infer that the the insider's portfolio has the structure $\pi_t^{\text{ins}} = \alpha_t + k_t^L$. He/she would then know k_t^L exactly. In the case when $L = \lambda W_1 + (1-\lambda)\epsilon$, where ϵ is an independent, standard normal, it is shown in [10] that the drift becomes:

$$k_t^x = \frac{(x - \lambda W_t)\lambda}{(\lambda^2)(1-s) + (1-\lambda)^2}$$

Again, if in addition we take the structure of L to be known, the value of the term x may be recovered. However it would be unreasonable to assume that the structure of L is known explicitly. In this case, the relationship between k_t^L and the distribution of L is complicated and cannot be inverted in general.

Nevertheless, the representation of the insider's information in (1.7) is appealing. One may relate the insider's optimization problem to that of the 'uniformed agent' by the following:

$$V^{\text{ins}}(x) = \sup_{\pi \in \mathcal{A}} \tilde{\mathbb{E}}\left[U(X_1^\pi)\right] = \sup_{\pi \in \mathcal{A}} \mathbb{E}\left[U(X_1^\pi)\frac{d\mathbb{P}}{d\tilde{\mathbb{P}}_1}\right] \tag{1.12}$$

where, $\tilde{\mathbb{P}}_1$ is the measure that makes $\tilde{W}_t$ a martingale. Instead of a logarithmic utility, if one where to consider an exponential utility, then setting

$C = \log\left(d\mathbb{P}/d\tilde{\mathbb{P}}_1\right)$ in the objective on the right hand side becomes:

$$\sup_{\pi \in \mathcal{A}} \mathbb{E}\left[U(X_1^\pi + C)\right] \tag{1.13}$$

In other words, the insider can be interpreted as maximizing utility in the presence of the contingent claim C, the logarithm of the Radon-Nikodym of the change of measure that makes $\tilde{W}$ a martingale. The problem of recovering the information can now be stated as trying to recover the structure of C, taking as input:

- The insider's optimal trading strategy π_t.
- The value of the prices P_t or, equivalently, the state variable W_t.

As such, some structure on C must be assumed to make this problem tractable. In this work, it is assumed that C is a function of the terminal value of the price P_T, that is, $C = g(P_T)$.

Such a problem is a novel type of inverse problem. Martin Klimmek has done interesting work regarding a different class of inverse problems in mathematical finance. In his PhD. thesis [7], he is concerned with finding diffusions that are consistent with the prices of perpetual American options. The prices of such options are given by $V(\theta) = \sup_\tau \mathbb{E}\left[e^{-\rho\tau} G(X_\tau, \theta)\right]$ and are parametrized by the strike θ. Klimmek approaches this problem using u-convex analysis. A function $f : D \to \mathbb{R}^+$ is u-convex if there exists non-empty sets $D_y \subset D$, $D_z \subset D$ and $S \subset D_z \times \mathbb{R}^+$, such that for all $y \in D_y$:

$$f(y) = \sup_{(z,a)\in S} \left[u(y,z) + a\right]$$

The u-dual of f is the u-convex function on D_z, given by:

$$f^u(z) = \sup_{y \in D_y} [u(y,z) - f(y)]$$

Klimmek is able to show that it is possible to find diffusion X that is consistent with a continuum of log process $v(\theta) = \log(V(\theta))$ with the following result:

Theorem 3. *There exists a diffusion X that is consistent with the continuum of prices $V(\theta)$ if and only if there exists $\phi : [0,\infty) \to [1,\infty)$ such that $\phi(0) = 1$, ϕ is increasing and convex and ϕ is such that v is the g-dual of $\log(\phi)$ where $g = \log(G(x,\theta))$.*

1.2 Utility Maximization

Consider the following setup. The financial market comprises of a stock S and a zero-interest bond. The dynamics of S are modeled by a geometric Brownian motion with constant drift μ and volatility $\sigma > 0$:

$$dS_t = S_t\Big(\mu\, dt + \sigma\, dB_t\Big)$$

A financial agent maximizes the expected value of the exponential utility function $U(x) = -e^{-\gamma x}$, applied to the combination of the gains from trading in the financial market and the random endowment of the form $C = g(S_T)$ (or equivalently, of the form, $C = \bar{g}(B_T)$). It is well-known, from [3], that his/her optimal portfolio takes the form

$$\hat{\pi}_t = \frac{\theta}{\sigma} - v_x(t, B_t),$$

where $\theta = \mu/\sigma$ and v uniquely (and classically) solves the linear parabolic Cauchy problem

$$\begin{cases} 0 = v_t + \frac{1}{2}v_{xx} - \theta v_x + \frac{1}{2\gamma}\theta^2 \\ v(T,x) = \bar{g}(x) \end{cases} \tag{1.14}$$

Hence, observing trajectories $\hat{\pi}_t$ and S_t of the agent's optimal investment strategy and of the risky asset is equivalent to observing the values of the solution of a linear parabolic Cauchy problem along a Brownian trajectory. After a straightforward change of variables, the parabolic equation (1.14) can be replaced by the heat equation with a minimally modified terminal condition, and we can focus on the following, linear, inverse problem:

Problem (L): Recover the unknown function g, given single trajectories W_t, and $\hat{P}_t = u(t, W_t)$, $t \in [0,T]$ of the Brownian motion and the solution u of the Cauchy problem

$$\begin{cases} 0 = u_t + \frac{1}{2}u_{xx} \\ u(T,x) = g(x) \end{cases} \tag{1.15}$$

for the heat equation, along W_t, $t \in [0,T]$.

Traditionally, inverse problems regarding the heat equation do not take this form. Much work has been done in solving the *backward heat equation*. This corresponds to recovering $g(x)$ knowing $u(x,t^*) = u_0(x), \quad \forall x$ for some fixed t^*. Without loss of generality, we take $t^* = 0$. Taking the Fourier transform of (1.15), we get

$$\begin{cases} 0 = \hat{u}_t - \frac{\xi^2}{2}\hat{u} \\ u(0,x) = \hat{u}_0 \end{cases} \tag{1.16}$$

the solution to which is $\hat{u} = e^{\frac{\xi^2}{2}t}\hat{u}_0$. Applying the inverse Fourier transform we get:

$$u(t,x) = \frac{1}{2\pi}\int e^{ix\xi+\frac{\xi^2}{2}t}\hat{u}_0(\xi)d\xi \tag{1.17}$$

Note that the term $e^{\frac{\xi^2}{2}t}$ in the integrand grows very fast t increases. Unless the function $\hat{u}$ grows faster than $O\left(e^{-x^2}\right)$ the integrand will diverge in finite time $\hat{t} > 0$. Consequently, the integral, and so the solution u, will not exist. This means that even a small perturbation in u_0 that does not decay faster than e^{-x^2} can dramatically distort the solution or even make the integral (1.17) infinite. This instability of the solution means that the problem is *ill-posed*.

One approach to overcome this ill-posedness is adding *regularity conditions* to the problem. There is a rich literature regarding this technique and its application, see [11], [4], [12]. The problem considered here is different. Instead of taking the solution $u(t^*, x)$ at fixed time t^* for all x as input we are taking the trajectory $\hat{P}_t = u(t, W_t)$. The solution (1.17) no longer applies and we are not able to rely on existing literature.

Another problem related to problem (L) arises by taking the values of u along a 'horizontal line' as input. That is, rather than taking the values $u(t^*, x)$ for a fixed t^* as input, one may consider taking the values $u(t, x^*)$ for a fixed x^* and $t \in [0, T]$ instead. Indeed, this resembles problem (L) more closely. Since it is possible to translate the problem by setting $y = x - x^*$, we take $x^* = 0$ for simplicity. Note that at any given time t, it is possible to interpret u, the solution to (1.15), as :

$$u(t,0) = \mathbb{E}\left[g(W_T)|W_t = x\right] = \frac{1}{\sqrt{2\pi\tau}}\int_{-\infty}^{\infty} g(x)e^{-\frac{x^2}{2\tau}}dx \tag{1.18}$$

where $\tau = T - t$. For convenience we write $u(0,t) = \phi(\tau)$ by simply reversing time. If we take $g(x)$ to be symmetric around 0 then we get:

$$\left(\frac{\pi\tau}{2}\right)^{\frac{1}{2}}\phi(\tau) = \int_0^{\infty} g(x)e^{-\frac{x^2}{2\tau}}dx \tag{1.19}$$

Since the domain of integration is positive, we may rewrite the integrand in terms of $\tilde{g}\left(\frac{x^2}{2}\right) = \frac{g(x)}{x}$. With a change of variables setting $y = \frac{x^2}{2}$ and relabeling $s = \frac{1}{\tau}$ we have:

$$F(s) = \left(\frac{\pi\tau}{2}\right)^{\frac{1}{2}}\phi\left(\frac{1}{s}\right) = \int_0^{\infty}\tilde{g}(y)e^{-ys}dx \tag{1.20}$$

In other words, it is possible, assuming g is symmetric, to rewrite the given input $u(t,0)$ for $t \in [0,T]$ as $F(s)$ for $s \in \left[\frac{1}{T},\infty\right)$. $F(s)$ has the form of a Laplace transform. Since we know $F(s)$ only on a part of $\mathbb{R}^+$, we may use Post-Widder inversion formula that states:

$$\tilde{g}(t) = \lim_{n\to\infty}\frac{(-1)^n}{n!}\left(\frac{n}{t}\right)^{n+1}F^{(n)}\left(\frac{n}{t}\right) \tag{1.21}$$

where $F^{(n)}$ is the $n^{(th)}$ derivative of F. Since the limit takes $n \to \infty$, note that the argument $\frac{n}{t}$ will eventually be in $\left[\frac{1}{T},\infty\right)$, the domain of the modified input. In this way, it is possible to recover the function g, under the assumption it is symmetric.

As in the case with the backward heat equation, problem (L) is essentially different than the 'horizontal line' case. Since the input $\hat{P}_t$ is given along the trajectory of a Brownian Motion W_t, the reduction (1.20) is not possible and we cannot apply the Post-Widder inversion formula. To solve problem (L), the present work relies on the structure inherited by $\hat{P}$ from the fact that it can be interpreted as a running conditional expectation of a function of a Brownian Motion at the terminal time.

Chapter 2

Solution To Problem L

Problem (L) admits many different solutions. One solution relies on the asymptotic behaviour of the Brownian Motion around 0. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the canonical Wiener space and let $W_t,\ t \in [0,T]$ be a Brownian Motion. Then by the law of iterated logarithms we have:

$$\limsup_{t \to 0^+} \frac{W_t}{\sqrt{2t \log\log\left(\frac{1}{t}\right)}} = 1 \quad \mathbb{P} - \text{a.s.} \tag{2.1}$$

Given the input of Problem (L), we may perform a change of variables for our convince. Since the path of W_t is given as input, we may define a new path $B_\tau = W_{T-\tau} - W_T$, for $\tau \in [0,T]$. Note that B_τ it is just time reversed Brownian Motion, and so it will also satisfy (2.1). Additionally, we may set $\hat{g}(x) = g(x + W_T)$. Essentially, the problem (L) has been rewritten as:

Problem ($\hat{L}$): Recover the unknown function $\hat{g}$, given single trajectories B_t, and $\hat{P}_t = u(t, B_t)$, $t \in [0,T]$ of the Brownian motion and the solution u of the Cauchy problem

$$\begin{cases} 0 = u_t - \frac{1}{2}u_{xx} \\ u(0,x) = \hat{g}(x) \end{cases} \tag{2.2}$$

For the moment we assume that $\hat{g}$ has a (locally uniformly convergent) Taylor expansion [1] and can be written as $\hat{g}(x) = \sum_{n=0}^{\infty} \hat{a}_n x^n$. To recover $\hat{g}$ it is sufficient to recover the coefficients $\hat{a}_n$. We do this recursively as follows:

- Note that $B_0 = 0$ and so we may set $\hat{g}(B_0) = \hat{a}_0$.
- Given $\{\hat{a}_k\}_{k \leq n}$ for any $n \geq 1$ we may set $\hat{g}_{n+1} = \hat{g} - \sum_{k=0}^{n} \hat{a}_k x^k = \sum_{k=n+1}^{\infty} \hat{a}_k x^k$. The term with the lowest power is $\hat{a}_{n+1} x^{n+1}$. Therefore we know that :

$$\limsup_{t \to 0^+} \frac{\hat{g}_{n+1}(B_t)}{\left(\sqrt{2t \log\log\left(\frac{1}{t}\right)}\right)^{n+1}} = \limsup_{t \to 0^+} \hat{a}_{n+1} \frac{B_t^{n+1}}{\left(\sqrt{2t \log\log\left(\frac{1}{t}\right)}\right)^{n+1}} = \hat{a}_{n+1} \tag{2.3}$$

  as the terms with higher powers of B_t will vanish.

We may then recover $g(x) = \hat{g}(x - W_T)$. Such an algorithm does fully recover the function g. In fact one only needs to know the $\hat{u}$ and $\hat{B}$ on $[0, t)$ for some $t > 0$. This corresponds to knowing the input to problem (L) for any arbitrarily small interval containing T. However, such a procedure is hard to implement practically since the lim sup computations in (2.3) are very sensitive to the measurement of B_t. Since the path B_t has infinite first-order variation, it would be preferable to avoid such a calculation. The algorithm described here avoids such unstable calculations. Before we describe the algorithm in detail, we consider a toy problem as motivation.

We construct a simplified version of Problem (L). Suppose that instead of the trajectory $\hat{P}_t$ we are instead given a smooth trajectory $\psi(t)$ and instead

[1] So that we may interchange limits and infinite sums

of solving a differential equation like (1.15) we are given that $\psi(t)$ is a polynomial. We are tasked with recovering the general form of ψ. More precisely we are given :

- The entire trajectory of the function $\psi : [0, T] \to \mathbb{R}$
- That the function $\psi(t) = \sum_{k=-0}^{N} a_k t^k$ is a polynomial of a known finite order N.

and charged with the task of recovering a_k. Even though $\psi(t)$ is smooth, since we want to generalize to $\hat{P}_t$, we want an algorithm that avoids setting coefficients to evaluations of $\psi(t)$. In this way, we avoid the instability that arises in the algorithm used to solve Problem $(\hat{L})$. With this constraint, we may recover the coefficients a_k that uniquely define ψ by differentiating as follows:

- Differentiate ψ exactly N times and set

  $$\hat{a}_N = \frac{1}{N!}\left(\frac{1}{T}\int_0^T \frac{d^N}{dt^N}\psi(t)dt\right)$$

  By integrating we avoid the undesirable evaluations described above.

- Given $\{a_k\}_{k=n+1}^{N}$ for any $n \leq N-1$ we set $\psi_n(t) = \psi(t) - \sum_{k=n+1}^{N} a_k t^k$. and then, differentiating exactly n times we set:

  $$\hat{a}_n = \frac{1}{n!}\left(\frac{1}{T}\int_0^T \frac{d^N}{dt^N}\psi_n(t)dt\right)$$

Note that we need not necessarily know the order N, we may simply perform this algorithm for increasingly large N till our recovered signal $\hat{p} = \sum_{k=0}^{N} \hat{a}_k t^k$ is the same as $\psi(t)$ or 'close' with respect to some norm. The goal is to use a similar algorithm on $\hat{P}_t = u(t, W_t)$, the input to problem (L). Unlike $\psi(t)$, however, the signal $\hat{P}_t$ does not have a Taylor expansion in t. Therefore we must find an analogous operation to differentiation with which to treat the signal. To do this we will be working with Hermite Polynomials.

2.1 Hermite Polynomials

The construction of Hermite polynomials and their desired properties as basis functions are a consequence of the Sturm-Liouville Theory. According to this theory it is possible to generate an orthogonal basis for weighted L^2 space by determining the eigenvalues and eigenfunctions of a differential operator, provided it satisfies some regularity conditions and does not have 0 as an eigenvalue. For our purposes, we will only consider functions that map $\mathbb{R}$ to $\mathbb{R}$, however the theory holds for more general mappings. For any strictly positive and continuous weight function w we define the inner product:

$$\langle f, g \rangle_w = \int_{\mathbb{R}} f(x)g(x)w(x)dx$$

Accordingly, the norm is defined by $||f||^2_{L^2(\mathbb{R}),w} = \langle f, f \rangle_w$ and the space $L^2_w(\mathbb{R})$ is the set of measurable functions for which $||f||^2_{L^2(\mathbb{R}),w} < \infty$.

The Hermite polynomials are generated by considering the weight function $w(x) = e^{-\frac{x^2}{2}}$ and the eigenvalue problem

$$\frac{d}{dx}\left(e^{-\frac{x^2}{2}}\frac{d}{dx}u\right) + \lambda e^{-\frac{x^2}{2}}u = 0 \tag{2.4}$$

The structure of the problem being $(pu')' + (\lambda w + q)u = 0$ is essential to Sturm-Liouville Theory. For us, $p = w = e^{-\frac{x^2}{2}}$ and $q = 0$. (2.4) is equivalent to the equation

$$u'' - xu' + \lambda u = 0 \tag{2.5}$$

The eigenvalues of this equation are given by $\lambda_n = n$ and the corresponding eigenfunctions, the Hermite Polynomials, are given by:

$$h_n(x) = \frac{(-1)^n}{n!}e^{\frac{1}{2}x^2}\frac{d^n}{dx^n}(e^{-\frac{1}{2}x^2}) \tag{2.6}$$

Or, more explicitly :

$$h_n(x) = \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k x^{n-2k}}{k!(n-2k)!2^k} \tag{2.7}$$

It is important to note that these h_n are orthogonal, but not ortho-normal. That is:

$$\int_{\mathbb{R}} h_n(x)h_m(x)w(x)dx = \frac{\sqrt{2\pi}}{n!}\delta_{nm}(x)$$

This normalization is chosen so that $\frac{d}{dx}h_n(x) = h_{n-1}(x)$, a relation that will make later computations more convenient. Consequently, the decomposition of any $g \in L^2_w(\mathbb{R})$ is $g = \sum a_n h_n$ where the a_n are given by:

$$a_n = \frac{n!}{\sqrt{2\pi}}\langle g, h_n\rangle \tag{2.8}$$

To make the importance of the construction of these basis functions clear, we state the following corollary:

Corollary 1. *Define the time dependent version of the Hermite polynomials as:*

$$H_n(t,x) = t^{\frac{n}{2}} h_n\left(\frac{x}{\sqrt{t}}\right) \tag{2.9}$$

for each n. Then, the function H_n is space-time harmonic. That is:

$$\frac{\partial}{\partial t} H_n + \frac{1}{2}\frac{\partial^2}{\partial x^2} H_n = 0 \tag{2.10}$$

Proof. This is derived by explicitly computing the derivatives on the left hand side of (2.10) to get:

$$\frac{1}{2} t^{\frac{n}{2}-1}\left(h''\left(\frac{x}{\sqrt{t}}\right) - \frac{x}{\sqrt{t}} h'\left(\frac{x}{\sqrt{t}}\right) + nh\left(\frac{x}{\sqrt{t}}\right)\right) = 0$$

since h_n solves (2.5) for the eigenvalue n. □

Take $(W_t, \mathcal{F}_t, \Omega)$ to be a standard Brownian Motion on a canonical Weiner space. Then, Itô's formula gives us:

$$H_n(t, W_t) = \int_0^t \left(\frac{\partial}{\partial s} H_n + \frac{1}{2}\frac{\partial^2}{\partial x^2} H_n\right) ds + \int_0^t \frac{\partial}{\partial x} H_n dW_s$$

Applying (2.10), we see that the the first integral is always zero and so $H_n(t, W_t)$ is a martingale. This property will play a key role in the construction of the algorithm below.

Finally we make a few useful definitions regarding the space $L^2_w(\mathbb{R})$. Since the functions h_n form a basis for $L^2_w(\mathbb{R})$, we say that a sequence of functions function $q_k = \sum a_n^{(k)} h_n$ converges to $g = \sum b_n h_n$ in the Gaussian L^2 norm if:

$$||g - q_k||_{L^2(\mathbb{R}),w} = \sum_{n=1}^{\infty} \int_{\mathbb{R}} (a_n^{(k)} - b_n)^2 h_n^2 w dx = \sqrt{2\pi} \sum_{n=1}^{\infty} \frac{(a_n - b_n)^2}{n!} \to 0$$

as $k \to \infty$. Next, since each function $g \in L^2_w(\mathbb{R})$ can be written as $g = \sum a_n h_n$ uniquely we may identify sequence $\{a_n\}_{n=0}^{\infty}$ with g. More precisely, define the set

$$\mathcal{A} = \{\{a_n\}_{n=0}^{\infty} \quad \text{s.t.} \quad \sum_n \frac{a_n^2}{n!} < \infty\} \tag{2.11}$$

Consider the map $\phi(g) \to \{a_n\}_{n=0}^{\infty}$ defined by $\phi(g) \to \{a_n\}_{n=0}^{\infty}$ where the coefficients a_n are given by (2.8). By construction of $\mathcal{A}$ and the uniqueness of the decomposition of each $g \in L^2(\mathbb{R})_w$, ϕ is a one to one mapping with $\phi^{-1}(\{a_n\}_{n=0}^{\infty}) \to g(x) = \sum a_n h_n(x)$. This bijection means that in order to identify g it is sufficient to identify each of the coefficients a_n.

2.2 Defining the Algorithm

The goal here is to construct a procedure that mimics that which was used in the toy example with polynomial $\psi(t)$. To motivate the construction, first we study how some properties of the time dependent versions of the Hermite polynomials, H_n, will be used. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a canonical Wiener space and let $(W_t, \mathcal{F}_t, \Omega)$ be a standard Brownian Motion with the 'usual conditions' on the filtration $\mathcal{F}_t$. Note that we may convert any problem with $T > 1$ to one with $T = 1$ by simply 'throwing out' the signal from $[0, T-1)$ and considering the problem starting at time $\hat{T} = T - 1$. For the case when time $T < 1$, we will see in a later section that our result will be unaffected. Hence, for the remainder of the work we will stay under the assumption:

Assumption 2. *For problem (L), we take the terminal value $T = 1$*

Additionally, since we will be working on a single trajectory of W_t, it is important to specify precisely what is meant by quadratic variation. Let $\mathcal{M}$ be the set of square integrable martingales, i.e. $\mathbb{E}[X_t^2] < \infty$ for $t \geq 0$. For any time $t \in [0,1]$ we specify the following sequence of partitions of $[0,t]$:

$$\Pi_n = \left\{ \frac{i}{2^n} t : 0 \leq i \leq 2^n \right\} \tag{2.12}$$

This sequence of partitions has the property their mesh $||\Pi_n|| = 2^{-n} \to 0$ as $n \to \infty$. For any $X \in \mathcal{M}$ we define the variation sum at time t

$$V_t(\Pi_n) = \sum_{k=1}^{2^n} |X_{t_k} - X_{t_{k-1}}|^2 \tag{2.13}$$

for each partition Π_n, where $t_k = kt/2^n$. From [6] we know that when $X \in \mathcal{M}$, the variations 2.13 converge to the quadratic variation $\langle X \rangle_t$. To ensure that our operations are well defined, we will restrict ourselves to $W \in \hat{\Omega} \subset \Omega$ such that

$$\lim_{n\to\infty} V_t(\Pi_n)(W_t) = \lim_{n\to\infty} \sum_{k=1}^{2^n} |W_{t_k} - W_{t_{k-1}}|^2 = t \tag{2.14}$$

in $\mathbb{L}^2$, with our selection partitions Π_n. Since these partitions π_n have a summable mesh size, Problem (2.9.8) in [6], guarantees that 2.14 will happen with probability 1, i.e., $\mathbb{P}(\hat{\Omega}) = 1$. In the probability 0 event that we are given a $W_t \notin \hat{\Omega}$, we do not offer a solution to Problem (L). However, if $W \in \hat{\Omega}$ the version of the quadratic variation defined by (2.14) coincides with the classical notion of quadratic variation.

As discussed in (2.1), for any $g \in L^2(\mathbb{R})_w$ we may write $g(x) = \sum_{n=0}^{\infty} a_n h_n(x)$.

Under the assumption $T = 1$ we will be able to write:

$$g(x) = \sum_{n=0}^{\infty} a_n h_n(x) = \sum_{n=0}^{\infty} a_n H_n(1, x) \tag{2.15}$$

Since the functions $H_n(t, x)$ are space-time Harmoic, note that we may write

$$\mathbb{E}\left[h_n(W_1)|\mathcal{F}_t\right] = \mathbb{E}\left[H_n(1, W_1)|\mathcal{F}_t\right] = H_n(t, W_t)$$

Therefore, we may write the signal $\hat{P}$ as:

$$\hat{P}_t = u(t, W_t) = \mathbb{E}\left[g(W_1)|\mathcal{F}_t\right] = \sum_{n=0}^{\infty} a_n H_n(t, W_t) \tag{2.16}$$

In this way, we may write the trajectory of the running conditional expectation $u(t, W_t) = \mathbb{E}\left[g(W_1)|\mathcal{F}_t\right]$ knowing only the coefficients a_n, the basis functions $H_n(t, x)$ and the trajectory W_t. Since the functions $H_n(t, x)$ are known, the only unknown that determines the signal $\hat{P}$ is the sequence $\{a_n\}_{n=0}^{\infty}$. Note that these are the same a_n that uniquely determine g.

To mimic the toy example, we would like to find an operation that, when applied to $H_n(t, W_t)$ yields $H_{n-1}(t, W_t)$. This operator would mimic the derivative that maps $t^n \longmapsto (n-1)t^{n-1}$. To this end, note that the Hermite polynomials satisfy the rule $h_n'(x) = h_{n-1}(x)$. In terms of H_n this gives us: $\frac{\partial}{\partial x} H_n(t, x) = H_{n-1}(t, x)$. Using Ito's formula, and the fact that H_n is space-time harmonic, we may rewrite $H_n(t, W_t)$ as :

$$H_n(t, W_t) = H_n(0, 0) + \int_0^t H_{n-1}(s, W_s) ds$$

Finally, taking the quadratic co-variation with respect to the driving Brownian motion W_t we have:

$$\langle H_t(\cdot, W), W \rangle_t = \int_0^t H_{n-1}(s, W_s) dW_s \tag{2.17}$$

Since the integrand on the right hand side is continuous, by the fundamental theorem of calculus, we arrive at:

$$\frac{d}{dt}\langle H_n(\cdot, W), W\rangle_t = H_{n-1}(t, W_t) \tag{2.18}$$

This is exactly the desired operation. Like in the case of t^n, we have taken a signal 'of order n' and converted it to a signal 'of order n-1.' To understand the use of this operation precisely, we begin by defining the operators that will be used in treating the observation.

From (2.1) we know that it is possible to identify any $g \in L^2(\mathbb{R})_w$ with the coefficients of it's Hermite expansion $\{a_n\}_{n=0}^{\infty}$. As in the beginning, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a canonical Wiener space and let $(W_t, \mathcal{F}_t, \Omega)$ be a standard Brownian Motion with the 'usual conditions' on the filtration $\mathcal{F}_t$. In anticipation of application of the operators yet to be defined, we restrict ourselves to the following subset of $L^2(\mathbb{R})_w$:

$$\mathcal{G} = \{g \in L^2(\mathbb{R})_w \text{ s.t. } g = \sum_n a_n h_n \text{ and } \sum_n |a_n| < \infty\} \tag{2.19}$$

Note that any truncation of g, $g_n = \sum_{k \geq n} a_k h_k$ is also an element of $\mathcal{G}$. In the following, as in (2.17) and (2.18), we will be fixing a particular path of W_t before we perform any operations. We know that with probability 1 any path $W \in \hat{\Omega}$ will be bounded. Let this bound be $M = M(W)$ such that $W_t < M$ on $[0, 1]$. We have the following Lemma:

Lemma 1. *Let M be the bound for W_t on $[0, 1]$. Then*

$$|H_n(t, W_t)| \leq e^M \quad \forall t \in [0, 1], \quad \forall n \geq 0 \tag{2.20}$$

In particular, we have that $|H_n(t,x)| \leq e^M$ on $[0,1] \times [-M,M]$

Proof. Recall the definition in (2.7). We then have:

$$|H_n(t,W_t)| = t^{n/2} \left| \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k \left(W_t/\sqrt{t}\right)^{n-2k}}{k!(n-2k)!2^k} \right| \leq \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{M^{n-2k}|\sqrt{t}|^{2k}}{(n-2k)!} \leq \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{M^{n-2k}}{(n-2k)!}$$

as $t \leq 1$. Note also that $\sum_{k=0}^{\lfloor n/2 \rfloor} \frac{M^{n-2k}}{(n-2k)!} \leq \sum_{k=0}^{\infty} \frac{M^k}{k!} = e^M$. □

Using this lemma, we see that for any $g \in \mathcal{G}$ and $W \in \hat{\Omega}$ fixed, for any $(x,t) \in [-M,M] \times [0,1]$ we have:

$$\sum_{n=l}^{\infty} |a_n H_{n-l}(s,s)| \leq e^M \sum_{n=0}^{\infty} |a_n| \tag{2.21}$$

Since $\frac{\partial}{\partial x} H_n(t,x) = H_{n-1}(t,x)$, this means that the series of derivatives in x is uniformly convergent on $[-M,M]$. Therefore the derivative of the sum is the sum of the derivatives. That is:

$$u_x(t,x) = \sum_{n=1}^{\infty} a_n H_{n-1}(t,x) \tag{2.22}$$

on $[0,1] \times (-M,M)$ In particular this gives us the following corollary:

Corollary 2. *For g, and W fixed as above we have:*

$$\langle u(\cdot,W), W \rangle_t = \sum_{n=0}^{\infty} a_n \langle H_n(\cdot,W), W \rangle_t \tag{2.23}$$

Proof. The proof is just an application of (2.23) and the fact that u solves

(1.15):

$$\begin{aligned}
\langle u(\cdot, W), W\rangle_t &= \int_0^t u_x(s, W_s)ds \\
&= \int_0^t \sum_{n=1}^{\infty} a_n H_{n-1}(s, W_s)ds = \sum_{n=1}^{\infty} a_n \int_0^t H_{n-1}(s, W_s)ds \\
&= \sum_{n=0}^{\infty} a_n \langle H_n(\cdot, W), W\rangle_t
\end{aligned}$$

Where we may switch the sum in the integral by Fubini's theorem. □

Additionally we would like to show that we can also take the derivative of (2.23) by passing the derivative underneath the sum. To this end we note that:

- That each $\int_0^t H_{n-1}(s, W_s)ds$ is differentiable in t by the fundamental theorem of calculus as each $H_{n-1}(s, W_s)$ is continuous.

- Taking the derivative we have:

$$\frac{d}{dt}\int_0^t H_k(s, W_s)ds = H_k(t, W_t)$$

  and so

$$\left|\sum_{n=1}^{\infty} a_n \int_0^t H_{n-1}(s, W_s)ds\right| = \sum_{n=1}^{\infty} |a_n H_{n-1}(t, W_t)| \leq e^M \sum_{n=0}^{\infty} |a_n|$$

  by Lemma 1.

Therefore, since we have sum of uniformly convergent derivatives we have the following corollary:

Corollary 3. *For g, and W fixed as above we have:*

$$\frac{d}{dt}\langle u(\cdot, W), W\rangle_t = \sum_{n=0}^{\infty} a_n \frac{d}{dt}\langle H_n(\cdot, W), W\rangle_t = \sum_{n=1}^{\infty} a_n H_{n-1}(t, W_t) \qquad (2.24)$$

What Corollaries 2 and 3 say is that we are allowed to take the quadratic variation of $g \in \mathcal{G}$ term by term and then take its derivative also term by term. To make use of these corollaries, during our operations, we want to stay within the realm of processes that may be expressed as

$$P_t = \sum_n a_n H(t, W_t)$$

To this end we define for each g, the set:

$$\mathcal{P}_g = \{\alpha_t : \alpha_t = \sum_{n=0}^{\infty} a_n H_n(W_t, t) \text{ for some } W \in \hat{\Omega}\} \qquad (2.25)$$

Next define:

$$\mathcal{P} = \bigcup_{g \in \mathcal{G}} \mathcal{P}_g \qquad (2.26)$$

We now define several operators; all of them implicitly depend on a (fixed) realization W_t, $t \in [0, 1]$ of a Brownian motion:

1. Let the **derivative operator** $\mathcal{C} : \mathcal{P} \to \mathcal{P}$; be defined as follows:

$$\mathcal{C}(P)_t = \frac{d}{dt}\langle P, W\rangle_t.$$

2. For each $g \in \mathcal{G}$ **observation operator** $\mathcal{A} : \mathcal{G} \to \mathcal{P}$ by

$$\mathcal{A}(g)_t = u(t, W_t) \text{ where } u \text{ solves (1.15) above.}$$

3. The **evaluation functional** $\mathcal{I} : \mathcal{P} \to \mathbb{R}$ is given by

$$\mathcal{I}(P) = \frac{1}{|A|} \int_A P_t \, dt \text{ where } A = \{t \leq 1 : W_t \leq \sqrt{t}\}$$

   The choice of A is technical and plays a role in computing estimates later.

4. Lastly, we define the n^{th}**-leading-coefficient functional** $\mathcal{K}_n : \mathcal{P} \to \mathbb{R}$ by

$$\mathcal{K}_n(P) = \mathcal{I}\mathcal{C}^n(P)$$

With our operators defined, we are now in a position to apply it to the solution of (1.15). Naturally, the structure of g is important. As with our toy example, we first examine how to deal with the case when the boundary condition g is a polynomial of a known order and then study how to extend this to a general, analytic, g.

2.3 Polynomial g

Suppose g is a polynomial of a known order n, and the given Brownian motion is W_t. We write it in the Hermite basis:

$$g(x) = \sum_{k=0}^{n} a_k h_k(x)$$

In this case the observation operator admits a particularly pleasant form:

$$\mathcal{A}(g)_t = \sum_{k=0}^{n} a_k H_k(t, W_t) \text{ where } H_k(t,x) = t^{k/2} h_k\left(\frac{x}{\sqrt{t}}\right)$$

Since $h'_k(x) = h_{k-1}(x)$, using Itô's formula and applying the derivative operator leads to a key reduction:

$$\mathcal{C}(A(g))_t = \sum_{k=1}^{n} a_k H_{k-1}(t, W_t)$$

An n-fold application of the derivative operator $\mathcal{C}$ yields that $\mathcal{C}^n(A(g))_t = a_n$ and so $\mathcal{K}_n(\mathcal{A}(g)) = a_n$. Since $\mathcal{K}_{n-1}(\mathcal{A}(g) - a_n H_n) = a_{n-1}$, etc., we have the following identities

$$a_n = \mathcal{K}_n(\mathcal{A}(g)) \tag{2.27}$$

$$a_{n-1} = \mathcal{K}_{n-1}(\mathcal{A}(g) - a_n H_n) \tag{2.28}$$

$$a_{n-2} = \mathcal{K}_{n-2}(\mathcal{A}(g) - a_n H_n - a_{n-1} H_{n-1}) \tag{2.29}$$

$$\dots$$

Given that $\mathcal{C}(H_k) = H_{k-1}$, defining $b_n = \mathcal{K}_n(\mathcal{A}(g))$ and rewriting $\mathcal{K}_n(H_{n-k}) = c_k$ produces the following this system of equations of Toeplitz type:

$$\begin{pmatrix} 1 & \dots & c_{n-2} & c_{n-1} & c_n \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & c_1 & c_2 \\ 0 & \dots & 0 & 1 & c_1 \\ 0 & \dots & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ \vdots \\ a_{n-2} \\ a_{n-1} \\ a_n \end{pmatrix} = \begin{pmatrix} b_0 \\ \vdots \\ b_{n-2} \\ b_{n-1} \\ b_n \end{pmatrix} \tag{2.30}$$

Though numerically ill conditioned in general, this system admits a unique solution. In this way, knowing that g a is polynomial of order n, given observation $\mathcal{A}(g)_t$ and the path of the Brownian motion W, we can recover the function g by observing the following procedure

- Compute the covariances and derivatives $\mathcal{C}^k(\mathcal{A}(g), W)_t$ for all k.

- Since g is a polynomial, there is a N such that $\mathcal{C}^k(\mathcal{A}(g), W)_t = 0$ for $k > N$

- Compute the matrix T by computing $\mathcal{IH}_k$ for all $k \leq N$

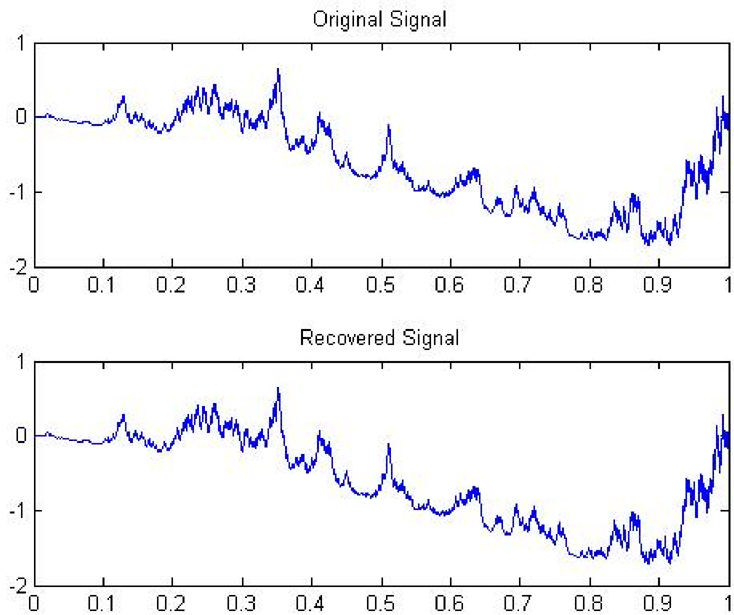
- Set a to be the coefficients obtained by $T^{-1}b$

This leads us to the following theorem:

Theorem 4. *Given a Brownian Motion W_t on $t \in [0,1]$, $\exists$ a sequence of invertible $n \times n$ linear operator T_n depending on the path of W_t such that:*

> *Given the observation $\mathcal{A}(g)_t$ for any polynomial function $g = \sum_{k=1}^{n} a_k h_k(x)$, the coefficients a_k can be recovered as the solution to the problem $a = T_n^{-1}b$, where $b_k = \mathcal{K}_k(\mathcal{A}(g))$ and $a(k) = a_k$*

Example: Set $g(x) = x^3 - 3x + x^2 - 1 = 6h_3(x) + 2h_2(x)$. Then our algorithm yields $\mathcal{C}^3(\mathcal{A}(g))_t \equiv 6$ and $\mathcal{K}_3(\mathcal{A}(g)) = 6$. Since $g - a_3h_3(x) = a_2h_2(x)$, we have $\mathcal{C}^2(\mathcal{A}(g - a_3h_3(x)))_t \equiv 2$ and $\mathcal{K}_2(\mathcal{A}(g - a_3h_3(x))) = 2$. A numerical implemented of the above on a simulated path yields the following results:



Top figure: the graph of the original signal $\mathcal{A}(g)$.
Bottom figure: the signal generated by the recovered function g.

Since the Brownian path was simulated by a random walk for this example, the set A of (3) for this example is just $[0, 1]$. Additionally, the quadratic covariation and derivative in the definition of the derivative operator $\mathcal{C}$ were approximated using finite differences. However, this naive implementation leads to some problems numerically. See discussion in the appendix.

2.4 General Functions

Though it is possible to recover polynomials of all order, it would be too optimistic to hope to be able to do the same for all functions $g \in L^2_w(\mathbb{R})$. When $g = \sum_{n=0}^{\infty} a_n h_n(x)$ is not a polynomial, we naturally approximate it by a polynomial by truncating the infinite sum. Then, we look for conditions on g for which the application of the algorithm converges as we truncate "later and later". Essentially, this amounts to studying the stability of the solutions $\{a_k\}_{k=1}^n$ as the size of the problem in (2.30) increases from n to $n+1$.

Since the terms c_k typically grow very fast, the system is inherently unstable. Indeed, one should expect fast decay in the the precision with which the finer and finer details of g are recovered. However, if one has an appropriate a-priori bound on the prominence of such fine details, i.e., if the terms $b_n = \mathcal{K}_n(\mathcal{A}(g))$ decrease fast enough, an increase in the size of the system should not disturb the "head" $\{a_k\}_{k=1}^n$ of the solution too much. Since the terms b_n satisfy

$$|b_n| = \frac{1}{|A|}\left|\int_A \sum_{l=1}^{\infty} a_{n+l} H_l(t, W_t)\, dt\right| \leq \sum_{l=1}^{\infty} |a_{n+l}|$$

the expression of interest is exactly the sum $\sum_{l=1}^{\infty} |a_{n+l}|$. For convininecе we call this sum $P = P(n) = \sum_{l=1}^{\infty} |a_{n+l}|$.

For a fixed n, solving the system 2.30 will generate approximations $\{\tilde{a}_{n-k}\}_{k=1}^n$. As per the algorithm, these approximations $\tilde{a}_{n-k}$ of a_{n-k} are given in (2.27 - 2.29). Since the observation operator $\mathcal{A}$ is linear we know that for any k:

$$\tilde{a}_{n-k-1} = \mathcal{K}_{n-k-1}(\mathcal{A}(\phi_k), W) \;\text{ where }\; \phi_k = g - \sum_{j=0}^{k} \tilde{a}_{n-j} H_{n-j}$$

From the definition of the evaluation operator, the approximation error of the highest order term being approximated will be:

$$|a_n - \tilde{a}_n| = \frac{1}{|A|}\left|\int_A \sum_{l=1}^{\infty} a_{n+l} H_l(t, W_t) dt\right|$$

To emulate the polynomial case, we seek a condition on a_n such that if we can make P small then $\tilde{a}_{n-k} - a_{n-k}$ would be small for all $k \leq n$. To this end, one has the following lemma:

Lemma 2. *For any given path of the Brownian motion W, there is a constant C, independent of n such that*

$$|H_n(t, W_t)| \leq t^{n/2} C \text{ on } \{|W_t| \leq \sqrt{t}\} \tag{2.31}$$

In particular, this means $\frac{1}{|A|}\left|\int_A H_l(t, W_t) dt\right| \leq C$ as $t \leq 1$.

Proof. Using the general expression for the Hermite polynomials we have:

$$|H_n(t, x)| = t^{n/2}\left|\sum_{k=0}^{n/2} \frac{-1^k\,(x)^{n-2k}}{k!(n-2k)!2^k}\right| \leq t^{n/2}\sum_{k=0}^{n/2} \frac{|x|^{n-2k}}{(n-2k)!}$$

Here, $|x| \leq 1$ as $|W_t| \leq \sqrt{t}$ and so the right hand side can be bounded above by $\sum_{k=0}^{\infty} \frac{1^m}{m!} = e$. The bound for the integral follows naturally.

□

Therefore, since $t \leq 1$, the error term is simply $|a_n - \tilde{a}_n| = C\sum_{l=1}^{\infty} |a_{n+l}| = CP$. The term $d_m = |a_m - \tilde{a}_m|$ can be written as follows:

$$d_{n-k} = \frac{1}{|A|} \int_{|A|} \left(\sum_{l=1}^{\infty} a_{n+l} H_{l+k} + \sum_{m=1}^{k} d_{n-k+m} H_m \right) dt \tag{2.32}$$

Where the first term corresponds to the error of truncation and the second term corresponds to error that arises from approximating the previous k coefficients $\{a_{n-j}\}_{j=1}^{k}$. Note that we accumulate more error as we iterate through our algorithm or, alternatively, as we solve (2.30) via Gaussian elimination. Naturally, we would want d_{n-k} to be small. We present the following lemma that relates $|d_{n-k}|$ to the tail term P:

Lemma 3. *For a fixed n, the error from following the program as in (2.27) is bounded by:*

$$|d_{n-k}| \leq C 2^k P \text{ for } k \leq n \tag{2.33}$$

Proof. For a fixed n, we proceed by induction on k. For $k = 1$ the difference is simply

$$|d_{n-1}| \leq \frac{1}{|A|} \int_{|A|} \sum_{l=1}^{\infty} |a_{n+l} H_{l+1}| dt + \frac{1}{|A|} \int_{|A|} |d_0 H_1| dt$$

Applying (2.31) and the fact that $|d_0| = P$ one has:

$$|d_{n-1}| \leq CP + CP = C2^1 P$$

Now, assuming the hypothesis for d_{n-m} for $m < k$, one again has:

$$\begin{aligned}
|d_{n-k}| &\leq \frac{1}{|A|}\int_{|A|}\sum_{l=1}^{\infty}|a_{n+l}H_{l+k}|dt + \frac{1}{|A|}\int_{|A|}\sum_{m=1}^{k}|d_{n-k+m}H_m|dt \\
&\leq CP + C\sum_{m=1}^{k}2^{k-m}P \\
&\leq CP\left(1+\sum_{j=0}^{k-1}2^j\right) = C\,2^kP
\end{aligned}$$

□

Note that the error can be re-written as $|d_m| \leq C2^{n-m}P$ for $m \leq n$. Therefore $(d_m)^2 \leq 2^{2(n-m)}C^2P^2$ for all $m \leq n$ as well as

$$\sum_{m=0}^{n}\frac{(d_m)^2}{m!} \leq C^2P^2\sum_{m=0}^{n}\frac{2^{2(n-m)}}{m!} \leq C^2\left(P2^n\right)^2\sum_{m=0}^{n}\frac{4^{-m}}{m!} \tag{2.34}$$

The infinite sum on the right is convergent and can be bounded above by $e^{1/4}$ for all n. We may therefore set $\hat{C} = e^{1/4}C^2$ which is independent of n. Therefore, for any n, we have the estimate:

$$\sum_{m=0}^{n}\frac{|d_m|^2}{m!} \leq \hat{C}\left(2^nP\right)^2 \tag{2.35}$$

where $P = \sum_{l=1}^{\infty}|a_{n+l}|$ is as defined above. We are now in a position to state the main result:

Theorem 5. *Given an observation $\mathcal{A}(g)_t$ (with $T = 1$), if the function g is such that*

$$2^n\sum_{l=1}^{\infty}|a_{n+l}| \to 0 \text{ as } n \to \infty \tag{2.36}$$

then the approximations of the n^{th} order $\tilde{g}_n = \sum_{k=0}^{n} \tilde{a}_k h_k$ generated from the procedure given by (2.27 - 2.29) converges to g in the Gaussian L^2 norm.

Proof. We take $\tilde{a}_k = 0$ for $k > n$ for notational convince. The norm of the difference $|g - \tilde{g}_n|_{L^2_w}$ induced by the Gaussian inner product is given by

$$
\begin{aligned}
|g - \tilde{g}_n|_{L^2(\mathbb{R}_w} &= \int_{\mathbb{R}} \sum_{m=0}^{\infty} (a_n - \tilde{a}_n)^2 h_n^2(x) w(x) dx = \sqrt{2\pi} \sum_{m=0}^{\infty} \frac{(a_n - \tilde{a}_n)^2}{m!} \\
&= \sqrt{2\pi} \left(\sum_{k=n+1}^{\infty} \frac{a_k^2}{k!} + \sum_{m=0}^{n} \frac{(d_m)^2}{m!} \right) \\
&\leq \sqrt{2\pi} \left(\sum_{k=n+1}^{\infty} \frac{a_k^2}{k!} + \hat{C} \left(2^n P\right)^2 \right)
\end{aligned}
$$

The first term, the sum, goes to zero as g is a L^2 function and the second does the same by the condition (2.36). In other words, for $\epsilon > 0$ we may take N to be the maximum such that for $n > N$ we have both

$$\sum_{k=n+1}^{\infty} \frac{a_k^2}{k!} < \frac{\epsilon}{2} \quad \text{and} \quad 2^n P \leq \left(\frac{\epsilon}{2\hat{C}} \right)^{\frac{1}{2}}$$

giving us:

$$|g - \tilde{g}_n|_{L^2_w} \leq \sqrt{2\pi}\epsilon$$

hence proving our claim. □

Remark 1. The assumption $T = 1$ is not crucial. We can write $g = \hat{g}(x/\sqrt{T}) = \sum_{n=0}^{\infty} a_n h(x/\sqrt{T})$. Now, setting $H_n(x,t) = h\left(\frac{x}{\sqrt{t}}\right) t^{n/2}$ we get that $H_n(W_t, t)$ is a martingale. The definition of H_n becomes:

$$H_n(t,x) = h\left(\frac{x}{\sqrt{t}}\right) \left(\frac{t}{T}\right)^{n/2}$$

We set the new leading coefficient operator $\hat{\mathcal{K}}_n = \mathcal{K}_n T^{n/2}$. Noting that the estimate in Lemma 1 will still hold (as there will be a $T^{n/2}$ in the denominator), the proofs will proceed unimpeded.

Chapter 3

The Quasilinear Problem

If we try to recover the agent's random endowment (or, equivalently, the form of his/her privileged information) in a more realistic, incomplete setting, we arrive to a nonlinear version of Problem (L). We present here only a single-variable caricature which nevertheless illustrates many of its salient features. For illustrative purposes, we also ignore the fact that equation we study can be turned into a heat equation using a simple exponential transform.

Problem (N): Recover the unknown function g, given single trajectories $\hat{B}_t$, and $u^\epsilon(t, \hat{B}_t),\ t \in [0, T]$ of the Brownian motion and the solution u of the quasilinear Cauchy problem

$$\begin{cases} 0 = u^\epsilon_t + \frac{1}{2} u^\epsilon_{xx} + \epsilon (u^\epsilon_x)^2 \\ u^\epsilon(T, x) = g(x) \end{cases}$$

along $\hat{B}_t$, $t \in [0, T]$.

We do not attempt to solve Problem (N) directly, but we study the quality of the approximation of its solution in the "small-ϵ-regime" by the solution of the linear Problem (L). We assume that g satisfies the conditions of Theorem 5 so that the algorithm for the linear problem described above is stable. Since the equations (2.27-2.29) are well-defined for $\mathcal{A}^\epsilon(g)_t = u^\epsilon(t, \hat{B}_t)$, we obtain the following system of equations:

$$\begin{pmatrix} 1 & \dots & c_{n-2} & c_{n-1} & c_n \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & c_1 & c_2 \\ 0 & \dots & 0 & 1 & c_1 \\ 0 & \dots & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_0^\epsilon \\ \vdots \\ a_{n-2}^\epsilon \\ a_{n-1}^\epsilon \\ a_n^\epsilon \end{pmatrix} = \begin{pmatrix} b_0^\epsilon \\ \vdots \\ b_{n-2}^\epsilon \\ b_{n-1}^\epsilon \\ b_n^\epsilon \end{pmatrix}$$

where $b_k^\epsilon = \mathcal{K}_k(\mathcal{A}^\epsilon(g)) = \mathcal{I}\mathcal{C}^k(\mathcal{A}^\epsilon(g))$, and the coefficients c_k remain unchanged.

We are now in a position to study the difference between the solutions a^ϵ to the system above and the solutions a of the system (2.30), corresponding to the linear case ($\epsilon = 0$). Recall that in the linear case we have $\mathcal{A}(g)_t = u(t, \hat{B}_t)$, where u satisfies the heat equation

$$u_t + \frac{1}{2} u_{xx} = 0 \qquad u(T, x) = g(x)$$

By construction we know that $\mathcal{C}^k(\mathcal{A}^\epsilon(g))_t = \frac{\partial^k}{\partial x^k} u^\epsilon(t, \hat{B}_t)$ so that

$$|b_k - \hat{b}_k| \leq |\tfrac{\partial^k}{\partial x^k}(u - u^\epsilon)|_{C^0}.$$

On the other hand, under the additional assumption that g is bounded, Schauder estimates and interpolation theorems for anisotropic Hölder spaces yield the following estimate: given $n \in \mathbb{N}$ and $\eta > 0$, there exists $\epsilon_0 > 0$ such that

$$\left| \tfrac{\partial^k}{\partial x^k}(u - u^\epsilon) \right|_{C^0} \leq \eta \text{ as soon as } k \leq n \text{ and } |\epsilon| \leq \epsilon_0.$$

For any fixed n, the Toeplitz matrix T, given in (2.30), is invertible and so $||T^{-1}||$, the matrix norm of T^{-1}, is finite. Hence, the difference $|a_k - a_k^\epsilon| \leq \eta ||T^{-1}||$ can be made small, provided ϵ is small enough:

Theorem 6. *Suppose g satisfies the hypothesis of Theorem 5 and is bounded, then for a_n^ϵ and a_n defined as above, $a_n^\epsilon \to a_n$, as $\epsilon \to 0$.*

In words, when ϵ is small, our algorithm for the linear Problem (L) produces a good approximation to the solution of the nonlinear problem (N). The proof of (6) is technical and requires application of Schauder estimates.

3.1 Proof of Theorem 6

We start by defining the two solution to the PDEs with which we are concerned.

$$h_t + \frac{1}{2}h_x x = 0 \qquad h(T,x) = g(x)$$

and

$$h_t^\epsilon + \frac{1}{2}h_x^\epsilon x + \epsilon(h_x^\epsilon)^2 = 0 \qquad h^\epsilon(T,x) = g(x)$$

Where the function g is bounded as above. We will follow the analysis of [8] in the following. We start with some definitions. Let $C(\mathbb{R} \times [0,T])$ be the set of all continuous functions $u : \mathbb{R} \times [0,T] \to \mathbb{R}$. Define the sup-norm as:

$$|u|_0 = \sup_{(x,t)\in\mathbb{R}\times[0,T]} |u(x,t)|$$

Next, for any function $u \in C(\mathbb{R} \times [0,T])$, define the α-Hölder constant, $\alpha \in (0,1]$, of $[u]_\alpha \in [0,\infty]$ by :

$$[u]_\alpha = \sup_{(x_1,t_1))\neq(x_2,t_2)} \frac{|(x_1,t_1) - u(x_2,t_2)|}{d_p\left((x_1,t_1) - (x_2,t_2)\right)^\alpha}$$

Where the parabolic distance $d_p\left((x_1,t_1) - (x_2,t_2)\right) = \sqrt{|t_2 - t_1|} + |x_2 - x_1|$. Furthermore, we define:

$$
\begin{aligned}
|u|_{\alpha} &= |u|_0 + [u]_{\alpha} \\
|u|_{1+\alpha} &= [u_x]_{\alpha} + |u|_0 + |u_x|_0 \\
|u|_{2+\alpha} &= [u_t]_{\alpha} + [u_{xx}]_{\alpha} + |u_t|_0 + |u_{xx}|_0 + |u|_0 + |u_x|_0
\end{aligned}
$$

The last expression is abbreviated as $|u|_{2+\alpha} = [u]_{2+\alpha} + |u_t|_0 + |u_{xx}|_0 + |u|_0 + |u_x|_0$. Additionally, we also use the following lemmas from [8].

Lemma 4. *There exists a universal constant $C > 0$ such that for any $\delta > 0$ and $u \in C^{\alpha+2}$ we have*

$$
\begin{aligned}
[u_x]_{\alpha} &\leq \delta[u]_{2+\alpha} + C\delta^{-(1+\alpha)}|u|_0 \\
[u_x]_0 &\leq \delta[u]_{2+\alpha} + C\delta^{-1/(1+\alpha)}|u|_0
\end{aligned}
$$

And another inequality:

Theorem 7. *There exists a universal constant $C > 0$ such that for any $u \in C^{\alpha+2}$ we have:*

$$
[u]_{2+\alpha} \leq C\left([u_t + \frac{1}{2}u_{xx}]_{\alpha} + [u(T,\cdot)]_{2+\alpha}\right)
$$

Lastly, we will be using the multiplicative form of Parabolic interpolation inequalities:

Theorem 8. *There exists a universal constant $C > 0$ such that for any $u \in C^{2+\alpha}$ we have:*

$$
\begin{aligned}
|u_x|_0^{2+\alpha} &\leq C\,[u]_{2+\alpha}\,|u|_0^{1+\alpha} \\
[u_x]_{\alpha}^{2+\alpha} &\leq C\,[u]_{2+\alpha}^{1+\alpha}\,|u|_0
\end{aligned}
$$

Lastly, We will use the shorthand $u \preceq v$ to mean that there exists a universal constant C such that $u \leq Cv$.

3.1.1 Computing estimates

We define $\delta = h^\epsilon - h$ and note that Δ satisfies the following equation:

$$\Delta_t + \frac{1}{2}\Delta_{xx} + \epsilon(\Delta_x - h_x)^2 = 0 \qquad \Delta(T, x) = 0$$

This can be re-written as:

$$\Delta_t + \frac{1}{2}\Delta_{xx} + \epsilon\left(\Delta_x^2 - 2\Delta_x h_x + h_x^2\right) = 0 \qquad \Delta(T, x) = 0$$

We note that by Feynman-Kac we have that

$$|\Delta|_0 \leq \epsilon |h_x^2|_0$$

Theorem 9. *Let Δ be as defined in (3.1.1). Then,* $[\Delta]_{2+\alpha} \to 0 \quad as \quad \epsilon \to 0$

Proof. By Theorem (7) we know that

$$[\Delta]_{2+\alpha} \leq C\epsilon\left[h_x^2 - 2h_x\Delta_x + \Delta_x^2\right]_\alpha \leq C\epsilon\left(\left[h_x^2\right]_\alpha + [2h_x\Delta_x]_\alpha + \left[\Delta_x^2\right]_\alpha\right) \quad (3.1)$$

Since h is the solution to the heat equation with bounded boundary condition we know, applying Theorem 7 and Theorem 8, that the first term in the sum is a constant. Using interpolation inequalities and the definition of $[\cdot]_\alpha$ we know that

$$\left[\Delta_x^2\right]_\alpha \leq 2|\Delta_x|_0\left[\Delta_x\right]_\alpha \leq 2\left[\Delta\right]_{2+\alpha}|\Delta|_0 \preceq \epsilon\left[\Delta\right]_{2+\alpha} \quad (3.2)$$

Similarly, for the second term we have

$$[h_x \Delta_x]_\alpha \leq |\Delta_x|_0 \, [h_x]_\alpha + |h_x|_0 \, [\Delta_x]_\alpha \preceq |\Delta_x|_0 + [\Delta_x]_\alpha$$

From the interpolation inequalities and the fact that ϵ can be taken to be less than 1 (and so :

$$\begin{aligned} |\Delta_x|_0 &\leq C \, [\Delta]_{2+\alpha}^{1/(2+\alpha)} K_1 \epsilon^{(1+\alpha)/(2+\alpha)} \preceq [\Delta]_{2+\alpha}^{1/(2+\alpha)} \epsilon^{1/(2+\alpha)} \\ [\Delta_x]_\alpha &\leq C \, [\Delta]_{2+\alpha}^{(1+\alpha)/(2+\alpha)} K_2 \epsilon^{1/(2+\alpha)} \end{aligned}$$

Combining this, and the above inequality we get that :

$$[h_x \Delta_x]_\alpha \preceq \epsilon^{1/(2+\alpha)} \left([\Delta]_{2+\alpha}^{1/(2+\alpha)} + [\Delta]_{2+\alpha}^{(1+\alpha)/(2+\alpha)} \right) \tag{3.3}$$

Combining (3.1), (3.2) and (3.3) we have

$$[\Delta]_{2+\alpha} \preceq \epsilon \left(1 + \epsilon \, [\Delta]_{2+\alpha} + \epsilon^{1/(2+\alpha)} \left([\Delta]_{2+\alpha}^{1/(2+\alpha)} + [\Delta]_{2+\alpha}^{(1+\alpha)/(2+\alpha)} \right) \right) \tag{3.4}$$

For any fixed ϵ, the value of $[\Delta]_{2+\alpha}$ is either greater than 1 or less than 1. If $[\Delta]_{2+\alpha} \leq 1$ then we have simply;

$$[\Delta]_{2+\alpha} \leq K\epsilon(1 + \epsilon + \epsilon^{1/(2+\alpha)}) \tag{3.5}$$

If, however, $[\Delta]_{2+\alpha} > 1$, because of concavity, $[\Delta]_{2+\alpha}^{1/(2+\alpha)}, [\Delta]_{2+\alpha}^{(1+\alpha)/(2+\alpha)} \leq [\Delta]_{2+\alpha}$. Thus, (3.4) reduces to

$$[\Delta]_{2+\alpha} \preceq \epsilon \left(1 + [\Delta]_{2+\alpha} \left(\epsilon + \epsilon^{1/(2+\alpha)} \right) \right)$$

In turn this implies:

$$[\Delta]_{2+\alpha} \leq \frac{K\epsilon}{1 - K\left(\epsilon + \epsilon^{1/(2+\alpha)}\right)} \tag{3.6}$$

Since the right hand side of both (3.5) and (3.6) go to zero as ϵ goes to zero, the desired conclusion follows $\square$

Lemma 5. *If* $u^\epsilon \in C^{2+\alpha}$*, and if* $[u^\epsilon]_{2+\alpha}$ *and* $[u^\epsilon]_0$ *tend to* 0 *as* $\epsilon \to 0$ *then,* $|u^\epsilon_x|_\alpha$ *also tends to* 0 *as* $\epsilon \to 0$

Proof. This follows simply by adding the two inequalities given by (4), for any fixed δ (we may fix $\delta = 1$). $\square$

Because of the above lemma and theorem we are able to say that $|\Delta_x|_\alpha \to 0$ as $\epsilon \to 0$. This will form the base case of the induction step below.

3.1.2 Induction

In this section we will be in need of the following lemma:

Lemma 6. *For any* $u \in C^{2+\alpha}$ *there is some universal constant* C *such that*

$$|u_x|_\alpha \leq C\left([u]_{2+\alpha} + |u|_0\right)$$

Proof. Using the additive interpolation inequalities we will have:

$$|u_x|_\alpha \leq \delta[u]_{2+\alpha} + C_1\left(\delta^{-(1+\alpha)} + \delta^{-1/(1+\alpha)}\right)|u|_0$$

For any fixed δ and for a universal C_1 (independent of δ). Setting C to be the maximum of the two coefficients, the desired conclusion follows. $\square$

We note that we may generate equations for the derivatives of Δ by differentiating (3.1.1) in x. The equations would involve the nth derivative of $(h_x^\epsilon)^2$. In general taking the n^{th} partial derivative of u^2 in x will yield 2^n terms of the form:

$$(u^2)^{(n)} = \sum_{k=0}^{n} C(n,k) u^{(n-k)} u^{(k)}$$

where we denote $u^{(n)}$ to mean the n^{th} partial derivative of u with respect to x and $C(n,k)$ to mean the combinatorial term 'n choose k'. We note that the only terms in this sum that involves the n^{th} derivative is the term $2u^{(n)}u$. Substituting $(\Delta_x - h_x)$ in for u we notice that this term can be expanded to get $\Delta_x^{(n)}(\Delta_x - h_x) - h_x^{(n)}(\Delta_x - h_x)$. Additionally, any other term is simply $(\Delta_x^{(n-k)} - h_x^{(n-k)})(\Delta_x^{(k)} - h_x^{(k)})$. If we assume that $|\Delta_x^{(k)}|_\alpha \to 0$ as $\epsilon \to 0$ for $k \leq n-1$ then we know that

$$\begin{aligned}
\left[(\Delta_x^{(n-k)} - h_x^{(n-k)})(\Delta_x^{(k)} - h_x^{(k)})\right]_\alpha &\leq 2\left|(\Delta_x^{(n-k)} - h_x^{(n-k)})\right|_\alpha \left|(\Delta_x^{(k)} - h_x^{(k)})\right|_\alpha \\
&\leq 2\left(|\Delta_x^{(n-k)}|_\alpha + |h_x^{(n-k)}|_\alpha\right)\left(|\Delta_x^{(k)}|_\alpha + |h_x^{(k)}|_\alpha\right) \\
&\leq 8|h_x^{(n-k)}|_\alpha |h_x^{(k)}|_\alpha
\end{aligned}$$

For ϵ small enough, for $k \leq n-1$. Indeed, for ϵ small enough we will also have $[h_x^{(n)}(\Delta_x - h_x)]_\alpha \leq 8|h_x^{(n)}|_\alpha |h_x|_\alpha$. In fact, though these ϵ many depend on k, we may choose ϵ small enough for all $k \leq n$. Since $|h_x^{(k)}|_\alpha$ are constants, indeed and bounded (uniformly) for $k \leq n$ (by the norm of the largest one) we can say that:

$$[\left((\Delta_x - h_x)^2\right)^{(n)}]_\alpha \leq M(n) + [\Delta_x^{(n)}(\Delta_x - h_x)]_\alpha$$

Where the constant $M(n)$, depends on n but not on ϵ, so long as ϵ is small enough. We may expand the second term to get:

$$\begin{aligned}
[\Delta_x^{(n)}(\Delta_x - h_x)]_\alpha &\leq [\Delta_x^{(n)}]_\alpha |(\Delta_x - h_x)|_\alpha + |(\Delta_x - h_x)|_\alpha |\Delta_x^{(n)}|_0 \\
&\leq |(\Delta_x - h_x)|_\alpha \left([\Delta_x^{(n)}]_\alpha + |\Delta_x^{(n)}|_0\right) \\
&\leq 2|h_x|_\alpha |\Delta_x^{(n)}|_\alpha \leq K_1 \left([\Delta^{(n)}]_{2+\alpha} + |\Delta_x^{(n-1)}|_0\right)
\end{aligned}$$

Where the last inequality follows from (6). Finally, we are able to state the following lemma:

Lemma 7. *If* $|\Delta_x^{(k)}|_\alpha \to 0$ *as* $\epsilon \to 0$ *for* $k \leq n-1$, *then*

$$[\left((\Delta_x - h_x)^2\right)^{(n)}]_\alpha \leq K(n)\left(1 + [\Delta^{(n)}]_{2+\alpha} + |\Delta_x^{n-1}|_0\right)$$

Proof. This follows directly from the above estimate on $[\Delta_x^{(n)}(\Delta_x - h_x)]_\alpha$ and (3.1.2) □

We now state and prove the induction step of the argument

Theorem 10. *If* $|\Delta_x^{(k)}|_\alpha \to 0$ *as* $\epsilon \to 0$ *for* $k \leq n-1$, *then* $|\Delta_x^{(n)}|_\alpha \to 0$ *as* $\epsilon \to 0$ *as well.*

Proof. By the (5) it is sufficient to show that $[\Delta^{(n)}]_{2+\alpha}$ goes to zero as $\epsilon \to 0$ as we already know $|\Delta^{(n)}|_0 = |\Delta_x^{(n-1)}|_0 \to 0$. We know that $\Delta^{(n)}$ satisfies the following equation:

$$\Delta_t^{(n)} + \frac{1}{2}\Delta_{xx}^{(n)} + \epsilon\frac{\partial^n}{\partial x^n}(\Delta_x - h_x)^2 = 0 \qquad \Delta(T,\cdot) = 0$$

As in the previous section, we are in the position to use (7) to get

$$[\Delta^{(n)}]_{2+\alpha} \leq C\epsilon[\big((\Delta_x - h_x)^2\big)^{(n)}]_\alpha$$

By (7) we get that

$$[\Delta^{(n)}]_{2+\alpha} \leq \epsilon\, CK(n)\left(1 + [\Delta^{(n)}]_{2+\alpha} + |\Delta_x^{n-1}|_0\right)$$

This implies

$$[\Delta^{(n)}]_{2+\alpha} \leq \frac{\epsilon CK(n)\,(1 + |\Delta_x^{n-1}|_0)}{1 - CK(n)\epsilon}$$

Again, as $\epsilon \to 0$ the above implies that $[\Delta^{(n)}]_{2+\alpha}$ goes to 0 (as $|\Delta_x^{n-1}|_0$ by the induction hypothesis) and the desired conclusion follows. □

Note that the induction for the statement that $|\Delta_x^{(n)}|_\alpha \to 0$ as $\epsilon \to 0$ is now complete, the base case being done in the previous section ($n = 0$). We are now able to state the following corollary

Corollary 4. *Fix n. For any $\eta > 0$ there exists an $\epsilon > 0$ such that*

$$|\Delta_x^{(k)}|_\alpha \leq \eta \qquad k \leq n$$

Proof. We know that for each k, there is an ϵ_k such that $|\Delta_x^{(k)}|_\alpha \leq \eta$. We obtain the desired result by setting $\epsilon = \min_{k \leq n} \epsilon_k$ □

Appendices

Appendix A

Numerical Implementation

The numerical implementation of the algorithm studied in this work, even on simulated trajectories and polynomial target functions, is problematic. For simplicity, we approximate the path W_t on $[0, 1]$ with a random walk. Dividing $[0, 1]$ into n equal length interval, and taking $\xi_1, \xi_2, ...\xi_n$ to be i.i.d. Bernoulli coin tosses with parameter $1/2$, the random walk X_t is given by $X_t = \frac{1}{\sqrt{n}} \sum_{k=1}^{t} \xi_k$. Suppose, as in the 'toy problem,' the goal was to recover the function g. The input to the algorithm would then be:

- The signal $\mathcal{A}(g)_t = \mathbb{E}\left[g(X_1)|X_t\right]$ for $t = \frac{k}{n}$ and $0 \leq k \leq n$
- The trajectory of the random walk X_t

We assume that the function g is a polynomial of a known order m. To solve the problem using the algorithm we must construct the matrix (2.30). This would, in particular, involve computing $\mathcal{C}^m(\mathcal{A}(g))_t$. The process of this would be as follows:

- Approximating the quadratic variation $\mathcal{C}(\mathcal{A}(g))_t = \frac{d}{dt}\langle \mathcal{A}(g), X\rangle_t$. We do this by setting:

$$\mathcal{C}(\mathcal{A}(g))_t \sim \frac{(\mathcal{A}(g)_{t+h} - \mathcal{A}(g)_t)\xi_t}{\xi_t^2} \tag{A.1}$$

- Then by computing $\mathcal{C}^k(\mathcal{A}(g))_t$ using the same approximation, setting

  $$\mathcal{C}^k(\mathcal{A}(g))_t \sim \frac{(\mathcal{C}^{k-1}(\mathcal{A}(g))_{t+h} - \mathcal{C}^{k-1}(\mathcal{A}(g))_t)\xi_t}{\xi_t^2}$$

  and repeating to compute $\mathcal{C}^m(\mathcal{A}(g))_t$

As per the algorithm, $\mathcal{K}(\mathcal{A}(g))$ should give us coefficient for the highest order Hermite polynomial.

Though the approximation (A.1) converges in the limit, the error terms in the approximations lead to problems. The trajectory $\mathcal{A}(g)_t$ should contain, among other terms, the term X_t^m as an approximation to W_t^m. This is the problematic as well as the relevant term. Setting $f(x) = x^m$ we may write:

$$f(X_t + \xi_t) = f(X_t) + f'(X_t)\xi_t + \frac{1}{2}f''(X_t)\xi_t^2 + ... \tag{A.2}$$

Our approximation then is a truncation of:

$$f'(X_t) = \frac{(f(X_t + \xi_t) - f(X_t))\xi_t}{h} - \frac{1}{2}f''(X_t)\xi_t + ...$$

Where $h = \frac{1}{n}$ as ξ_t^2 will always be positive. Call the term $\frac{(f(X_t+\xi_t)-f(X_t))\xi_t}{h} = \Delta_t$ for simplicity. The difference

$$f'(X_t + \xi_t) - f'(X_t) = \Delta_{t+1} - \Delta_t - \frac{1}{2}\left(f''(X_{t+1})\xi_{t+1} - f''(X_t)\xi_t\right) + .. \tag{A.3}$$

Our approximation would then set $f''(X_t) \sim \frac{(\Delta_{t+1}-\Delta_t)\xi_t}{h}$. That is:

$$f''(X_t) \sim \frac{(\Delta_{t+1} - \Delta_t)\xi_t}{h} - \frac{1}{2}(f''(X_{t+1})\text{sign}(\xi_{t+1}\xi_t) - f''(X_t))$$

Consider the difference in the expression on the right hand side. We may re write this as:

$$f''(X_t + \xi_t)\text{sign}(\xi_{t+1}\xi_t) - f''(X_t)$$

If $\text{sign}(\xi_{t+1}\xi_t)$ is positive, then the difference can be approximated by $f'''(X_t)\xi_t$. This would be desirable, as then the first order of the error term is of order $|\xi_k| \sim \frac{1}{\sqrt{n}}$. However, ξ_t and ξ_{t+1} are independent and so the sign could also be negative. In this case we simply have:

$$\frac{(\Delta_{t+1} - \Delta_t)\xi_t}{h} \sim f''(X_t) - \frac{1}{2}(f''(X_{t+1}) + f''(X_t))$$

The right hand side is clearly not the desired $f''(X_t)$. Immediately, one may be tempted to fix this issue by looking for a section of the random walk $X_t \ldots X_{t+m}$ have increments of the same sign, that is, such that $\{\xi_{t+k}\}_k^{k+m}$ all have the same sign. This way, we would never have $\text{sign}(\xi_{t+1}\xi_t) = -1$ and we stay away from the above complication. This is not a satisfactory fix not only because it doesn't necessarily extend to a different simulation of the Brownian path, but also because it is extremely sensitive the measurement of X_t. Alternatively, one may just keep track of the unwanted term $\frac{1}{2}(f''(X_{t+1}) + f''(X_t))$ and other terms like it that arise from repeated application of the derivative operator. This is possible if the target g is a lower order polynomial. However, this is extremely tedious for higher order polynomials. Moreover, this 'keeping track' method fundamentally changes the algorithm. Neither of these fixes are natural, however are the ones that are practically implementable.

A.1 Providing a Fix

Providing an effective numerical implementation that overcomes this problem requires further investigation. Since the algorithm behaves well when the target polynomial is of lower order, that is done to display the algorithm. A theoretical implementation can be provided but is numerically implemented.

Note that the approximations (A.1) is accurate in the limit as the step size $h = \xi_i^2$ goes to zero. For a target function of order m, we may construct m nested (uniform) partitions $\{P_i\}_{i=1}^m$ of $[0,1]$. That is $P_{k+1} \subset P_k$. We say that the step size of partition P_k is δ_k. We then simulate the random walk on the finest of the partitions $P_m : X_t = \sqrt{\delta_m}\sum_{k=1}^{1/\delta_m} \xi_k$. Note that the 'step size' for partition P_k would be δ_k. So the difference $X_{t+\delta_p} - X_t = \sqrt{\delta_m}\sum_{k=t}^{t+\delta_p} \xi_k$. This is not the same as above. Indeed, $|X_{t+\delta_p} - X_t| \neq \sqrt{\delta_p}$ in general.

The idea is to use increasingly coarser partitions for different evaluations of the derivative operator. As above, we compute

$$\mathcal{C}(\mathcal{A}(g))_t \sim \frac{(\mathcal{A}(g)_{t+\delta_m} - \mathcal{A}(g)_t)\xi_t}{\xi_t^2}$$

However, instead of using $\frac{(\Delta_{t+1}-\Delta_t)\xi_t}{\delta_m}$ as an approximation to $f''(X_t)$, we use

$$\frac{(\Delta_{t+\delta_{m-1}} - \Delta_t)(\xi_{t+\delta_{m-1}} - \xi_t)}{(\xi_{t+\delta_{m-1}} - \xi_t)^2}$$

In this case the right hand side of (A.2) becomes:

$$f''(X_t) - \frac{1}{2}\left(f''(X_t + \xi_t)\mathrm{sign}(\xi_{t+1}\xi_t) - f''(X_t)\right)\frac{\sqrt{\delta_m}}{(\xi_{t+\delta_{m-1}} - \xi_t)}$$

Intuitively, what we have done is 'taken the limit' using the partition P_m and then are using this on partition P_{m-1}. We will then use this as a limit for the stage P_{m-2}. Note that $|(\xi_{t+\delta_{m-1}} - \xi_t)| \sim \sqrt{\delta_{m-1}}$ and so the ratio

$$\frac{\sqrt{\delta_m}}{(\xi_{t+\delta_{m-1}} - \xi_t)} \sim \left(\frac{\delta_m}{\delta_{m-1}}\right)^{\frac{1}{2}}$$

Note that this is not exact, however is no worse than the approximation above. The order of this term is controlled by the ratio of partition sizes δ_m and δ_{m-1}. Naturally, we may construct our partitions P_k such that $\delta_m \ll \delta_{m-1}$ so that, say $\left(\frac{\delta_m}{\delta_{m-1}}\right)^{\frac{1}{2}} = \epsilon$. This way, we would expect our approximation

$$\frac{(\Delta_{t+\delta_{m-1}} - \Delta_t)(\xi_{t+\delta_{m-1}} - \xi_t)}{(\xi_{t+\delta_{m-1}} - \xi_t)^2} \sim f''(X_t) + o(\epsilon) \tag{A.4}$$

Furthermore, we may construct all partitions P_k such that $\left(\frac{\delta_{k+1}}{\delta_k}\right)^{\frac{1}{2}} = \epsilon$ so that each subsequent step will have the same order error.

Though robust, this method also has shortcomings. First, it is too computationally expensive. If we would want $\epsilon \sim 1/100$, we would need $\delta_{k+1} = \frac{\delta_k}{100^2}$ and, therefore $\delta_m = \frac{1}{100^{2m}}$. This type of resolution is unreasonable. Second, it isn't always the case that the difference $|(\xi_{t+\delta_{m-1}} - \xi_t)| \sim \sqrt{\delta_{m-1}}$. In fact, $\mathbb{P}(|(\xi_{t+\delta_k} - \xi_t)| \geq c\sqrt{\delta_k}$ for some $0 < c < 1$ is well known and converges as the partition size δ_{m-1} goes to zero. Hence, again, to make it so we can take multiple derivatives, we must only take sequences $\xi_t....\xi_{t+M}$ such that $|(\xi_{t+\delta_k} - \xi_t)| \geq c\sqrt{\delta_k}$ for each k. Though, this resembles the selection of the Brownian path mentioned above, this has a nicer interpretation. Essentially, controlled by the constant c, we are looking for a section of the random walk X_t that 'doesn't stay flat.'.

Bibliography

[1] Jürgen Amendinger, Peter Imkeller, and Martin Schweizer. Additional logarithmic utility of an insider. *Stochastic Process. Appl.*, 75(2):263–286, 1998.

[2] Kerry Back. Insider trading in continuous time. *The Review of Financial Studies*, 5(3):387–409, 1992.

[3] Freddy Delbaen, Peter Grandits, Thorsten Rheinländer, Dominick Samperi, Martin Schweizer, and Christophe Stricker. Exponential hedging and entropic penalties. *Math. Finance*, 12(2):99–123, 2002.

[4] M. Denche and A. Abdessemed. On regularization and error estimates for non-homogeneous backward Cauchy problem. *Arab J. Math. Sci.*, 18(2):149–164, 2012.

[5] J. Jacod. Grossisement initial, hopothese (h'), et theoreme de girsanov. *Grossisements de Filtrations: Exemples et Applications.*, 1985.

[6] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[7] M. Klimmek. *On inverse problems in mathematical finance.* PhD thesis, University of Warwick, 2012.

[8] N. V. Krylov. *Lectures on elliptic and parabolic equations in Hölder spaces*, volume 12 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1996.

[9] Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

[10] Igor Pikovsky and Ioannis Karatzas. Anticipative portfolio optimization. *Adv. in Appl. Probab.*, 28(4):1095–1122, 1996.

[11] Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of ill-posed problems*. V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York-Toronto, Ont.-London, 1977. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.

[12] Nguyen Huy Tuan, Dang Duc Trong, and Pham Hoang Quan. On a backward Cauchy problem associated with continuous spectrum operator. *Nonlinear Anal.*, 73(7):1966–1972, 2010.